



THỐNG KÊ TOÁN HỌC

LÝ THUYẾT MẪU

Sunday, May 9, 2021

1

THỐNG KÊ TOÁN HỌC

KIỂM ĐỊNH

Monday, May 12, 2025

1

Kiểm định giả thiết thống kê

1 Khái niệm

2 Kiểm định giá trị tham số

- Kiểm định giả thiết về kì vọng
- Kiểm định giả thiết về tỷ lệ

3 Kiểm định so sánh tham số

- Kiểm định so sánh giá trị trung bình
- Kiểm định so sánh tỷ lệ

Giả thiết và đối thiết

Bước đầu tiên của bài toán kiểm định là phát biểu chính xác điều sẽ kiểm định, đó là **giả thiết và đối thiết**.

H_0 : giả thiết (the null hypothesis)

H_1 : đối thiết (the alternative hypothesis)

Kết quả của mỗi kiểm định là hoặc chấp nhận H_0 hoặc bác bỏ (từ chối) H_0 và chấp nhận H_1 .

VD: Để xác định tốc độ kết nối trung bình có phải là 54 Mbps, ta kiểm tra giả thiết:

$$H_0 : \mu = 54, \quad H_1 : \mu \neq 54$$

Giả thiết và đối thiết

Chú ý: Giả thiết H_0 luôn được phát biểu dưới dạng đẳng thức.

$$H_0 : \mu = \mu_0$$

Đối thiết H_1 là một trong các dạng sau:

$$H_1 : \mu \neq \mu_0$$

$$H_1 : \mu < \mu_0$$

$$H_1 : \mu > \mu_0$$

Sai lầm trong kiểm định và mức ý nghĩa

Sai lầm loại I: H_0 đúng nhưng bác bỏ H_0 .

Sai lầm loại II: H_0 sai nhưng chấp nhận H_0 .

Trong thực tế, sai lầm loại I được xem là nguy hiểm hơn, do đó ta sẽ thiết kế mô hình kiểm định sao cho xác suất sai lầm loại I xảy ra bị chặn bởi một số rất nhỏ α .

Định nghĩa 1

α được gọi là **mức ý nghĩa** của kiểm định nếu

$$\alpha = P(\text{bác bỏ } H_0 | H_0 \text{ đúng})$$

Phương pháp tổng quát

Kiểm định giả thiết được dựa vào một thống kê kiểm định T , là đại lượng được tính từ dữ liệu thu thập được.

Nếu H_0 đúng thì T sẽ tuân theo một phân phối nào đó. VD: Trong bài toán kiểm định giá trị trung bình, thì T thường có phân phối chuẩn tắc hoặc pp Student.

Giả sử R là một miền giá trị thực sao cho

$$P(T \in R | H_0) = \alpha$$

với α rất bé, nghĩa là trong điều kiện H_0 đúng, khả năng xảy ra $T \in R$ gần như bằng 0.

Nếu từ dữ liệu ta có $T \in R$ thì ta bác bỏ giả thiết H_0 .

Nếu từ dữ liệu ta có $T \notin R$, như vậy không có cơ sở để bác bỏ H_0 , do đó ta vẫn chấp nhận H_0 .

$\Rightarrow R$: miền bác bỏ H_0 .

Kiểm định giả thiết về kì vọng

Giả sử X có phân phối chuẩn $N(\mu, \sigma^2)$.

Gọi μ là giá trị trung bình của tổng thể. Ta sẽ kiểm định giả thiết

$$H_0 : \mu = \mu_0$$

Giả sử H_0 đúng, nghĩa là $\mu = \mu_0$, thì

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Kí hiệu $z_{\alpha/2}$ là giá trị thỏa

$$P(|Z| > z_{\alpha/2}) = \alpha \iff P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

Kí hiệu z_α là giá trị thỏa

$$P(Z > z_\alpha) = \alpha \iff P(Z \leq z_\alpha) = 1 - \alpha$$

TH: σ đã biết (X có phân phối chuẩn)

(Hoặc X có phân phối bất kì nhưng $n \geq 30$)

Đặt

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

- **Bài toán 1:** $H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$
Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
Chấp nhận H_0 nếu $|z| \leq z_{\alpha/2}$
- **Bài toán 2:** $H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$
Bác bỏ H_0 nếu $z > z_{\alpha}$
Chấp nhận H_0 nếu $z \leq z_{\alpha}$
- **Bài toán 3:** $H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0$
Bác bỏ H_0 nếu $z < -z_{\alpha}$
Chấp nhận H_0 nếu $z \geq -z_{\alpha}$

TH: σ chưa biết

a) $n \geq 30$

Đặt

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

với s là độ lệch chuẩn có hiệu chỉnh của BNN X tính từ mẫu dữ liệu.

- **Bài toán 1:** $H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$
Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
Chấp nhận H_0 nếu $|z| \leq z_{\alpha/2}$
- **Bài toán 2:** $H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$
Bác bỏ H_0 nếu $z > z_{\alpha}$
Chấp nhận H_0 nếu $z \leq z_{\alpha}$
- **Bài toán 3:** $H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0$
Bác bỏ H_0 nếu $z < -z_{\alpha}$
Chấp nhận H_0 nếu $z \geq -z_{\alpha}$

TH: σ chưa biết

b) $n < 30$, X có phân phối chuẩn

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim \text{PP Student bậc tự do } n - 1$$

với s là độ lệch chuẩn có hiệu chỉnh của BNN X tính từ mẫu dữ liệu.

Kí hiệu t_α và $t_{\alpha/2}$ là các số thực thỏa

$$P(T > t_\alpha) = \alpha, \quad P(T > t_{\alpha/2}) = \frac{\alpha}{2}$$

với T tuân theo phân phối Student bậc tự do $n - 1$ (tra bảng A5)

TH: σ chưa biết, X có phân phối chuẩn

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

b) $n < 30$

- **Bài toán 1:** $H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$
Bác bỏ H_0 nếu $|t| > t_{\alpha/2}$
Chấp nhận H_0 nếu $|t| \leq t_{\alpha/2}$
- **Bài toán 2:** $H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$
Bác bỏ H_0 nếu $t > t_{\alpha}$
Chấp nhận H_0 nếu $t \leq t_{\alpha}$
- **Bài toán 3:** $H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0$
Bác bỏ H_0 nếu $t < -t_{\alpha}$
Chấp nhận H_0 nếu $t \geq -t_{\alpha}$

Kiểm định giả thiết về tỷ lệ

Giả sử ta quan tâm đến những phần tử có đặc trưng A nào đó trong tổng thể.

Gọi p là tỷ lệ phần tử có đặc trưng A trong tổng thể.

Ta sẽ kiểm định giả thiết $H_0 : p = p_0$.

Xét thống kê (với n lớn)

$$Z = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} \sim N(0, 1)$$

Kiểm định giả thiết về tỷ lệ

Dựa trên mẫu dữ liệu, tính

$$z = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$$

với f là tỷ lệ phần tử có tính chất A trong mẫu cụ thể.

- **Bài toán 1:** $H_0 : p = p_0, \quad H_1 : p \neq p_0$
Bác bỏ H_0 nếu $|z| > z_{\alpha/2}$
Chấp nhận H_0 nếu $|z| \leq z_{\alpha/2}$
- **Bài toán 2:** $H_0 : p = p_0, \quad H_1 : p > p_0$
Bác bỏ H_0 nếu $z > z_{\alpha}$
Chấp nhận H_0 nếu $z \leq z_{\alpha}$
- **Bài toán 3:** $H_0 : p = p_0, \quad H_1 : p < p_0$
Bác bỏ H_0 nếu $z < -z_{\alpha}$
Chấp nhận H_0 nếu $z \geq -z_{\alpha}$

Ví dụ 1.

Một hãng buôn muốn biết xem phải chăng có sự không ổn định về lượng hàng bán được trung bình trên một nhân viên bán hàng so với các năm trước (lượng đó bằng 7,4). Một mẫu ngẫu nhiên gồm 40 nhân viên bán hàng được lựa chọn và tìm thấy lượng hàng trung bình của họ là $\bar{x} = 6,1$ với độ lệch chuẩn có hiệu chỉnh là $s=2,5$. Với mức ý nghĩa $\alpha = 1\%$ có thể nói rằng lượng hàng bán được trung bình trên mỗi đầu người có sự thay đổi không?

Ví dụ 2.

Một công ty có một hệ thống máy tính có thể xử lý 1200 hóa đơn trong một giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lý trung bình trong một giờ là 1260 với độ lệch chuẩn có hiệu chỉnh là 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Ví dụ 3.

Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình là 21,5 tạ/ha. Gieo thử hạt giống mới này tại 16 vườn thí nghiệm và thu được kết quả:

19,2; 18,7; 22,4; 20,3; 16,8; 25,1; 17,0; 15,8; 21,0; 18,6;
23,7; 24,1; 23,4; 19,8; 21,7; 18,9.

Dựa vào kết quả này hãy xác nhận xem quảng cáo của công ty có đúng không? Mức ý nghĩa được lựa chọn là $\alpha = 0,05$. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân phối chuẩn $N(\mu, \sigma^2)$.

Ví dụ 4.

Một đảng chính trị trong một cuộc bầu cử tổng thống ở nước họ tuyên bố rằng có 45% cử tri sẽ bỏ phiếu cho ứng cử viên A của đảng họ.

Chọn ngẫu nhiên 2000 cử tri để thăm dò ý kiến và cho thấy có 862 cử tri tuyên bố sẽ bỏ phiếu cho ứng cử viên A.

Với mức $\alpha = 5\%$, hãy kiểm định xem dự đoán của đảng trên có đúng không.

BÀI TẬP

5.27 Trọng lượng đóng bao của một loại sản phẩm X là biến ngẫu nhiên có phân bố theo quy luật chuẩn với trọng lượng trung bình theo quy định là 100kg. Nghi ngờ sản phẩm bị đóng thiếu, người ta cân thử 29 bao loại này ta thu được kết quả:

Trọng lượng (kg)	98,0 - 98,5	98,5 - 99,0	99,0 - 99,5	99,5 - 100	100 - 100,5	100,5 - 101
Số bao tương ứng	2	6	10	7	3	1

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

5.28 Định mức thời gian hoàn thành sản phẩm là 14 phút. Liệu có cần thay đổi định mức không, nếu theo dõi thời gian hoàn thành sản phẩm ở 250 công nhân ta thu được kết quả như sau:

X (phút)	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
Số công nhân	20	60	100	40	30

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý định nói trên.

BÀI TẬP

5.29 Mức hao phí xăng của một loại ô tô chạy từ A đến B là một biến ngẫu nhiên có quy luật chuẩn với kỳ vọng 50 lít. Đoạn đường được sửa chữa lại. Người ta cho rằng mức hao phí xăng trung bình giảm xuống. Quan sát 28 ô tô cùng loại thu được

X hao phí (lít)	48,5 - 49,0	49,0 - 49,5	49,5 - 50,0	50,0 - 50,5	50,5-51
Số ô tô tương ứng	4	10	9	3	2

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

5.30 Một công ty có một hệ thống máy tính có thể xử lý 1300 hoá đơn trong 1 giờ. Công ty mới nhập một hệ thống máy tính mới, hệ thống này chạy kiểm tra trong 40 giờ cho thấy số hoá đơn xử lý trung bình trong 1 giờ là 1378 với độ lệch tiêu chuẩn 215. Với mức ý nghĩa 2,5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?